Timbre blending of wind instruments: acoustics and perception

Sven-Amin Lembke
CIRMMT / Music Technology
Schulich School of Music, McGill University
sven-amin.lembke@mail.mcgill.ca

Stephen McAdams
CIRMMT / Music Technology
Schulich School of Music, McGill University
smc@music.mcgill.ca

ABSTRACT

The acoustical and perceptual factors involved in timbre blending between orchestral wind instruments are investigated based on a pitch-invariant acoustical description of wind instruments. This description involves the estimation of spectral envelopes and identification of prominent spectral maxima or 'formants'. A possible perceptual relevance for these formants is tested in two experiments employing different behavioral tasks. Relative frequency location and magnitude differences between formants can be shown to bear a pitch-invariant perceptual relevance to blend for several instruments, with these findings contributing to a perceptual theory of orchestration.

Keywords

timbre perception, blend, orchestration, auditory fusion, spectral envelope

1. BACKGROUND

Timbre blending between instruments is a common application in orchestration practice. Important perceptual cues for blend are known to be based on note onset synchrony or partial tone harmonicity [5], which rely mainly on rhythmic or pitch relationships and hence compositional and performance factors. An orchestrator's choice of instruments, on the other hand, is more likely motivated by acoustical features of particular instruments.

Previous studies have suggested the perceptual relevance of pitch-invariant spectral traits characterizing the timbre of orchestral wind instruments. Analogous to human voice formants, the existence of stable local spectral maxima across a wide pitch range has been reported for these instruments [7, 3]. Furthermore, coincidence of these formant regions between instrumental sounds has been argued to contribute to the percept of blend between timbres [4].

Our aim is to verify and validate these hypotheses based on a two-stage approach comprising acoustical description and perceptual investigation. An attempt is made to correlate instrument usage with acoustical and perceptual factors by investigating whether a perceptual relevance of pitch-invariant spectral traits can be shown.

2. ACOUSTICAL DESCRIPTION

Spectral analyses are computed on a broad audio sample database across the entire pitch range of instruments. Based on the obtained spectral information, partial tones are identified and their frequencies and amplitudes used to build global distributions of partials across all available pitches and between dynamic levels. A curve-fitting procedure applied to these empirically derived distributions yields a spectral envelope estimate from which pitch-invariant traits such as formant regions are identified and described, as shown in Figure 2.
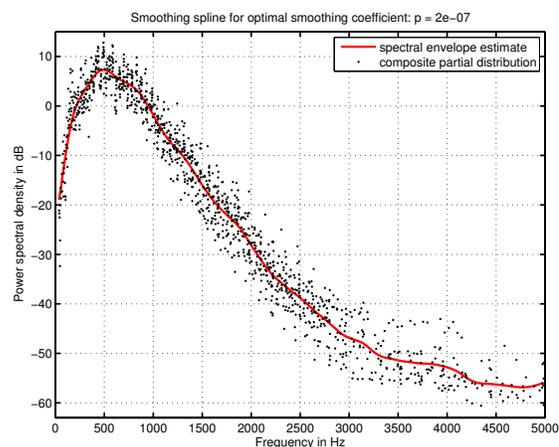

Figure 1. Spectral envelope estimate for bass trombone (line) and distribution of partial tones (dots).

As a means to investigate the perceptual relevance of the spectral traits a sound synthesis model is designed, based on two independently controllable formant filters with their frequency responses matched to the spectral envelope estimates. The synthesis is incorporated into a stimulus-presentation environment allowing real-time spectral shape modification, with its sound forming a dyad with a recorded wind instrument sound. The spectral shape modifications were operationalized as deviations for two formant-filter parameters for each formant i: 1) formant frequency ΔF_i in Hz and 2) formant magnitude ΔL_i in dB. The zero-deviation case represents the so-called *ideal* which corresponds to the originally modelled filter frequency response.

3. PERCEPTUAL INVESTIGATION

3.1 Experimental design

The perceptual relevance was assessed through two behavioral experiments that differed in the experimental tasks, with the second also aiming to provide further validation and clarification of findings from the first experiment. The synthesized instruments were paired with recorded samples of the same instruments at selected pitches. The instruments investigated in the main experiments were bassoon, horn, trumpet, oboe, flute and clarinet. Besides providing a validation for the contribution of formant regions to perceptual blend for different instruments, the experiment's multifactorial design also allowed their relevance to be investigated across different pitches, intervals and registers. With respect to multifactorial statistical hypothesis tests, both experiments adopted a within-participants design.[1]

3.1.1 Experiment A: blend production

The first experiment employed a production task and was conducted with 17 participants, recruited as musically experienced listeners. Across 88 trials (22 conditions × 4 repetitions) participants were given the task to adjust either ΔF_i or ΔL_i directly in order to achieve the maximum attainable blend. User control of the stimulus production environment was provided via a two-dimensional graphical interface, with controls for the investigated formant parameter and the loudness balance between instruments. The parameter deviations from ideal values were taken as the dependent variable.

3.1.2 Experiment B: blend rating

The second experiment was based on a simplified and less time-consuming rating task and involved 20 participants, again recruited as experienced listeners. Across 120 trials (30 conditions × 2 contexts × 2 repetitions) participants

were asked to rate the relative degree of blend for a total of 5 sound dyads per condition. A continuous relative blend rating scale was used, spanning from *most blended* to *least blended*. Across 5 dyads the same instrument sample formed pairs with varying formant parameter value presets for ΔF_1 or ΔL_1, with only the main formant ($i = 1$) being considered.[2] For both parameters one of the presets presented the zero-deviation *ideal* case. The remaining 4 presets comprised moderate deviations below (-*mod*) and above (+*mod*) the ideal and likewise, a pair of extreme deviations (-*ext* and +*ext*). These presets were based on generalizable formant properties which allowed comparisons between instruments to be made on a common scale of spectral-envelope description.

For ΔF_1 the 4 non-ideal preset values were defined as formant-frequency deviations corresponding to the points at which the spectral-envelope magnitude had decreased by either 3 dB (mod) or 6 dB (ext) relative to the formant maximum (see Figure 2). For ΔL_1 the moderate deviations represented values obtained from the behavioral findings of Experiment A paired with values mirrored relative to the ideal. The extreme deviations were defined as being 60% more extreme than the moderate ones.
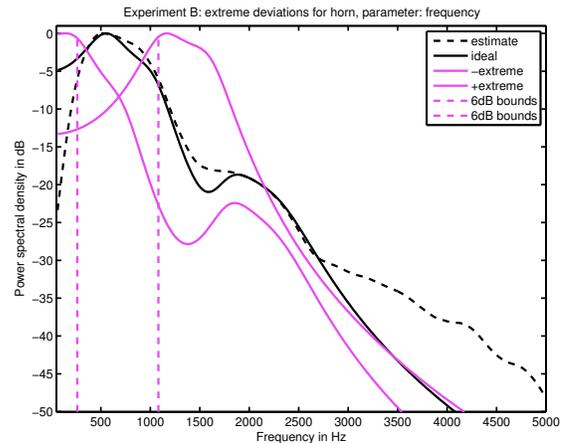
**Figure 2. Extreme deviations based on 6 dB bounds.**

Perceptual performance for each instrument was assessed across 2-4 pitches to investigate whether rating trends for the parameter presets were stable across pitch. Furthermore, the 4 repetitions of each experimental condition included two contextual versions, involving the omission of either the preset for negative or positive extreme deviation which allowed us to assess whether contextual variations affected rating trends.

---

[1] All reported statistically significant results are based on a significance level: $\alpha = .05$.

[2] The presets included predetermined values for the loudness balance between instruments and also had been equalized for loudness across presets.

3.2 Behavioral findings

Experiment A yielded results for the scenario in which participants themselves determined the parameter values leading to the best perceived blend. For relative parameter deviations $\Delta F_1/f_{max}$ (normalized to the frequency of the formant maximum), a common trend to slightly underestimate the ideal by about 10% was found, as shown in Figure 3. For 4 instruments, the underestimations were statistically significant ($t(16) \leq -3.83, p \leq .0015, \eta \geq .692$), determined through a single-sample t-test against a sample mean of zero. Notably, the horn and bassoon did not differ significantly from the ideal formant frequency. The absolute deviations ΔL_1 showed a clear trend to relative amplification of the main formant contributing to best blend, results for all considered instruments (bassoon, horn, oboe) being significantly different from the ideal ($t(16) \geq 7.33$, $p < .0001$, $\eta > .870$).

Relative formant deviation
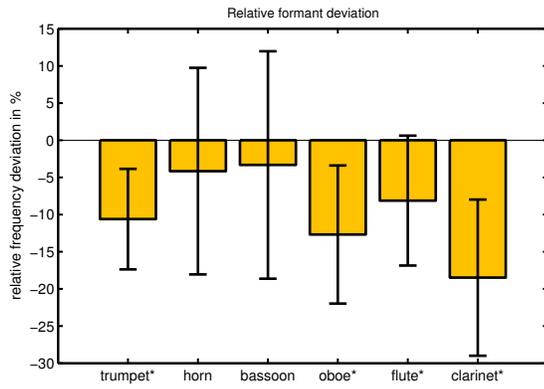
Figure 3. Mean behavioral $\Delta F_1/f_{max}$ (error: std. dev.).

No consistent significant trends can be reported for instruments compared across 3 interval types (unison, and non-unison consonance and dissonance) in a one-way ANOVA. Notably, across all tested instruments no indication was obtained that consonance or dissonance affected the chosen location of ΔF_1 differently. Another comparison between low versus high instrument register yields strong significant effects for all compared instruments (trumpet and bassoon: $F(1, 16) \geq 19.20$, $p \leq .0005, \eta_p^2 \geq .545$; clarinet: $F(1, 16) = 5.25, p = .0358, \eta_p^2 = .247$), suggesting that the perceptual relevance of formants does not hold at high registers. This finding was anticipated given the acoustical explanation that at high pitches the increased sparsity of partial tones outline formants inadequately, rendering them less meaningful as perceptual cues.

Experiment B aimed to confirm tendencies found in Experiment A and investigate whether they exhibited pitch invariance across a set of representative pitches, including the original conditions from the previous experiment. Instead of finding the best blend along a continuum of parameter deviations as in Experiment A, participants compared the relative degree of perceived blend between presets, which could, and in fact did, lead to some differences in the results. With regard to frequency deviations ΔF_1, the preferred (i.e. highest-rated) presets were not only oriented toward the ideal value and moderate underestimations (-*mod*), but included the extreme underestimation (-*ext*) as well. Conversely, the lowest ratings were obtained for overestimations of the ideal value (+*mod* and +*ext*), which agrees with the general trend of underestimation found in Experiment A. For gain deviations ΔL_i, amplification of the main formant could again be confirmed for the same instruments as in Experiment A, with nearly all comparisons being significantly different from the ideal. However, the trumpet, which had not been tested in Experiment A, did not show a clear trend for main formant amplification.

Several ANOVAs were conducted to investigate whether the findings argue for robust perceptual performance of ΔF_1-ratings across pitches, intervals and contexts.[3] The analysis rationale involved showing main effects for the factor 'preset' which would confirm that ratings could be considered as a reliable indicator for perceptual differences. Furthermore, the finding of significant interaction effects between the factors 'preset × pitch' would argue against pitch invariance, as the profile of blend ratings across presets would be shown to vary as a function of pitch. Likewise, obtaining interaction effects 'preset × interval' would reveal a different perceptual performance across unison and non-unison dissonance conditions. Finally, testing for main effects for 'context' assesses the robustness of perceptual findings across variations of stimulus context, for which only the presets common to both contexts, namely the *ideal* and the moderate deviations, are taken into account and normalized to the same scale limits.

Strong main effects for 'preset' are found across all instruments, indicating their utility to be taken as a measure of perceptual performance. Based on the multifactorial tests the 6 instruments form two groups. The grouping is based on whether or not significant deviations against the assumption of pitch-invariant perceptual relevance have been found, more specifically concerning significant interactions with either pitch or interval across both contexts or main

---

[3]Due to violations of the assumption of normality for about half the presets, main and interaction effects were tested with a battery of 5 independent ANOVAs on the raw and transformed behavioral ratings, including non-parametric approaches of rank-transformation [1] and prior alignment of 'nuissance' factors [2]. The most liberal and conservative p-values are reported. Whenever statistical significance is in doubt, the most conservative finding is assumed valid.

effects between contexts. Only statistically significant effects are reported below, with statistics taken from multiple ANOVAs reported as follows, e.g. *statistic = conservative value | liberal value*, and 'low' and 'high' denoting the contexts.

3.2.1 Group 1: pitch-variant

The pitch-variant group consists of flute and clarinet. The flute yields moderate interaction effects with pitch across both contexts (low: $F(3.95, 74.97 | 6, 114) = 2.83 | 4.29, p = .0311 | .0006, \eta_p^2 = .130 | .184$; high: $F(6, 114) = 4.41 | 7.19, p = .0005 | < .0001, \eta_p^2 = .188 | .275$) as well as a main effect for contextual variation ($F(1, 19) = 4.90 | 15.70, p = .0393 | .0008, \eta_p^2 = .205 | .452$). The clarinet exhibits a significant interaction effect across intervals for both contexts (low: $F(3, 57) = 2.88 | 3.38, p = .0440 | .0244, \eta_p^2 = .131 | .151$; high: $F(2.03, 38.66 | 3, 57) = 3.83 | 4.38, p = .0297 | .0076, \eta_p^2 = .168 | .188$). Although no significant interaction with pitch is obtained for the most conservative statistic, it should be noted that the most liberal findings for clarinet across both contexts (low: $F(3, 57) = 6.20, p = .0010, \eta_p^2 = .246$; high: $F(2.21, 41.92) = 19.86, p < .0001, \eta_p^2 = .511$) indicate even stronger effects than obtained for flute. As a result, flute and clarinet deliver clear indications for a departure from the assumption of pitch-invariant perceptual relevance. Interestingly, they also present the instruments that are the least-well represented by the acoustical formant description.

3.2.2 Group 2: pitch-invariant

Pitch-invariant perceptual relevance based on the formant description can be assumed for horn, bassoon, oboe and trumpet, given that no clear and consistent deviations from stable perceptual performance across pitch, interval and context were found. Among this group, the trumpet appears the least robust, as for non-unison interval type a single main effect for 'context' was obtained ($F(1, 19) = 10.50 | 25.04, p = .0043 | < .0001, \eta_p^2 = .355 | .569$). Attempting an acoustical explanation, this could possibly be explained by its acoustical description exhibiting a very broad formant which may not function to the same extent as the narrower and more defined main formants as found for the other three instruments.

Although Experiments A and B display somewhat different results concerning the perceptual relevance of exact overlap of the formants, they both support the hypothesis that perceived blend is achieved around and below the ideal formant location and is clearly reduced above this value. To further elucidate this tendency across all pitch-invariant instruments, a cluster analysis was conducted with the rating differences between preset levels being interpreted as a dissimilarity measure. This measure considered effect sizes (r)

of statistically significant non-parametric post-hoc analyses for pairwise comparisons between presets (Wilcoxon signed-rank test).[4] The *complete*-linkage clustering algorithm considered dissimilarity data averaged across 30 independent sets of effect sizes for the 4 instruments. As shown in Figure 4, the overestimations of ΔF_1 (+*mod* and +*ext*) are maximally dissimilar to a compact cluster associating deviations centered on and below the ideal formant location (*ideal*, -*mod* and -*ext*).
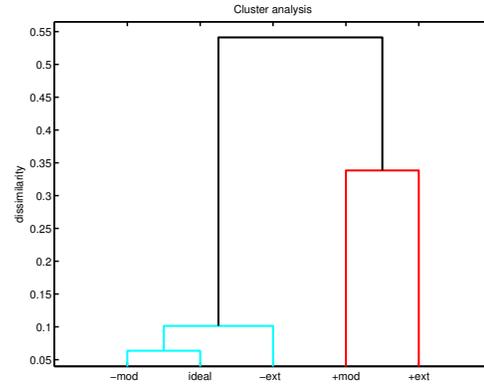

Figure 4. Dendrogram displaying clustering based on effect size from post-hoc analyses for 'preset'.

4. CONCLUSION

We have shown that localized formant regions are perceptually relevant to blend for the main formant parameters describing relative magnitude and frequency location. With regard to the former, a preference of relative main formant amplification could inversely be interpreted as a general attenuation of higher spectral-envelope traits. This can be taken as an implication that higher degrees of timbre blending may generally be achieved at lower dynamic markings (e.g., *mf*, *p*, *pp*), as it has been shown that secondary formants are less pronounced at lower excitation intensities [7].

As concerns the role of relative frequency location, the theory of formant coincidence [4] does not appear to hold across both investigated experimental tasks. Instead, it becomes clearly apparent that the role of formants in the perception of blend may function as a critical frequency boundary. The degree of perceived blend decreases markedly whenever the relative location of formants exceeds the frequency boundary of a *reference* formant (see Figure 5). As the reference formant in our investigation was predetermined by the static sampled instrument, it remains to be studied how this would apply to musical practice, in which musicians perform blend in an interactive relationship.

---

[4] Dissimilarity was assumed to be zero for non-significant differences.
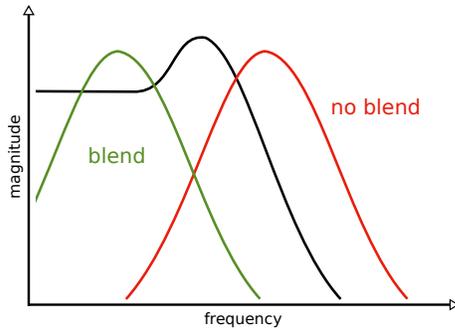
Figure 5. Schematic of theory of perceptual blend based on formant frequency relationships.

Pitch invariance is suggested by both the acoustical description and perceptual findings for most of the investigated wind instruments, which bears important implications for musical and orchestration practice. As the link between acoustical relationships and their contribution to perceptual blend has been established, pitch-invariant descriptors describing the frequency boundary may be able to serve as acoustical predictors of perceived blend. For the pitch-invariant instruments, this would enable the generation of systematic tables for blend relationships between combinations of different instruments and dynamic markings, which would serve as a helpful tool to orchestration practitioners. Furthermore, pitch invariance also suggests the utility of extending the notion of blend to non-unison usage in melodic-coupling or chordal phrases, as we have obtained clear findings arguing for a perceptual indifference to interval and consonance type. The single limitation of applicability concerns the perceptual relevance likely being unwarranted at the highest instrument registers as found in Experiment A, as would also be expected by acoustical considerations.

In general, our behavioral findings suggest that instruments' perceptual performance is most stable when strong formant cues are available acoustically, i.e. at pitch ranges that yield higher quantities and densities of partial tones to outline the spectral-envelope traits. This is in agreement with our findings obtained for the bassoon and horn, which in Experiment B exhibited a notable robustness across pitch and in Experiment A led to behavioral blend preference corresponding to the ideal formant location. Apart from being commonly used in orchestration practice to achieve blend, their lower pitch ranges could furthermore support a hypothesis of 'darker' timbres generally leading to more blend [6]. With this hypothesis having been derived from an acoustic description based on a global spectral average (e.g., spectral centroid), our investigation has contributed further by delivering more differentiated explanations based on a more local spectral origin.

These conclusions are expected to aid in the establishment of a spectral theory for perceptual blend that would serve as an instrument-specific complement to the composition- or performance-related cues mentioned in the introduction. A general perceptual theory for timbre blending could thereupon serve as a basis for reviewing existing treatises on orchestration concerning their agreement with the perceptual realities. It could also inspire new approaches to contemporary orchestration practice. At this point it can be hypothesized that rules established for formant-characterized instruments may concern a subset of possible perceptual blend scenarios. Given that they concern important members of the wind instrument family and in orchestration practice these are commonly given special care and attention, they might assume a critical role in a generalized theory of perceptual blend.

5. ACKNOWLEDGMENTS

REFERENCES

[1] Conover, W. J., and Iman, R. L. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician 35*, 3 (1981), 124–129.

[2] Higgins, J. J., and Tashtoush, S. An aligned rank transform test for interaction. *Nonlinear World 1* (1994), 201–221.

[3] Luce, D., and Clark, J. Physical correlates of brass-instrument tones. *The Journal of the Acoustical Society of America 42*, 6 (1967), 1232–1243.

[4] Reuter, C. *Die auditive Diskrimination von Orchesterinstrumenten - Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel.* Peter Lang, Frankfurt am Main, 1996.

[5] Sandell, G. J. *Concurrent timbres in orchestration: a perceptual study of factors determining blend.* Doctoral dissertation, Northwestern University, 1991.

[6] Sandell, G. J. Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception 13* (1995), 209–246.

[7] Schumann, K. E. *Physik der Klangfarben - Vol. 2.* Professorial dissertation, Universität Berlin, 1929.